

Ghost Voters?

An Abridged
Technical Review of the
2018 Zimbabwe's
Harmonised Election Voters' Roll

A report by Team Pachedu
Twitter: @pachedu2018

16 July 2018

Executive Summary

This report presents a concise analysis of the 2018 Voters' Roll, that complements the other report, the main aim being to address one main question: Are there ghosts in the voters' roll?

Several techniques were used to help explore this question, in light of the emerging evidence from the magnitude of inconsistencies that we originally uncovered (as detailed in Report 1). To achieve this end, we deployed computational tools and techniques to sift through the records in the voters' roll. Two main approaches were used in this regard viz. external consistency evaluation along with internal consistency evaluation.

The first approach was from a demographic perspective. The main finding that emerged was that the demographic distribution in the 2018 Voters' Roll is not consistent with the prevailing demographic models for Zimbabwe, along with precedent models from the 1982 up to the 2012 censuses. The main argument derived was that of possible manipulation of the entries on the voters' roll.

The second approach, which entailed the evaluation of the cases in the voters' roll was mainly anchored on confirming or disconfirming the coherence in the entries. To help put the 2018 entries into perspective, previous records were used, and these comprised of the 2008 and 2013 voters' rolls. One key finding was that for virtually all the quasi-similar ID entries from the 2018 roll, one of the pairs had an audit trail from prior voters' rolls, while the other did not, thereby reinforcing the argument that virtually all these cases were indeed duplicated cases, with slight alterations to the ID suffixes. Other inconsistencies were uncovered, with the key ones appearing to implicate the Registrar's Office owing to the emerging possibility of deliberate efforts to alter IDs, unless there are other genuine explanations for the discrepancies.

The major conclusion is that, by all standards, the 2018 Voters' Roll is fraught with errors and omissions, along with instances that lead to the assessment that 'ghost voters' do exist, despite the fact that ZEC had availed itself of advanced biometric registration and deduplication systems.

Table of Contents

1. Introduction	1
2. Methodology	1
3. Results	2
3.1 Demographic Distribution	2
3.2 External Inconsistency as Evidence of Potential Manipulation	5
3.3 Internal Inconsistency as Evidence of Potential Manipulation	21
4. Conclusion	31

1. Introduction

The initial intention of Team Pachedu's analysis was to evaluate the extent of internal consistency, and thus credibility that could be identified within the Zimbabwean 2018 Elections Voters' Roll. However, some of the anomalies identified in the first report prompted further inquiry, from a technical point of view, in an attempt to help establish whether the incidences were mere coincidence or, indeed, provided evidence of tampering. This phenomenon is generally known in popular parlance as 'ghost voters,' and in legal parlance, illegal voters. While there is not an agreed term to best define a 'ghost voter', in the context of this evaluation, the definition of a ghost voter was taken to imply a rogue entry, or a possible illegal voter, registered to vote in the 2018 election, by virtue of the voter being present more than once, or under unusual circumstances in the Zimbabwean Voters' roll released on the 18th of June 2018 by ZEC.

Based on this definition, this analysis identified ghost voters as comprising, but not limited to, duplicate entries, in part and in full, along with hierarchical incongruity with prior registration records, based on the 2013 and 2008 voters' rolls, along with other uncommon cases. It was, nonetheless, beyond the scope of this evaluation to then probe further through in-person visits and/or to investigate whether the flagged suspected ghosts would be used fraudulently or not as a means to cast illegal votes. This was the major limitation owing to the legal and logistical challenges entailed. However, the key goal of the evaluation was to help pinpoint all the key discrepancies and inaccuracies in the voters' roll, and it was envisaged that the responsibility to further validate the flagged individuals was left to the AFIS system. With this in mind, the ultimate goal of the study was to extract all suspect registrations, if there were any. The latter were, nonetheless, validated as existing, based on several exploratory and confirmatory criteria that shall be unfolded herein.

2. Methodology

To help achieve the task of sifting through the millions of entries in the voters' roll in search of any possible rogue entries, several computational and statistical techniques were used. Principally, computational techniques were used for the archival, extraction, transformation, loading and analysis of the outputs and this was facilitated by the use of custom-made Python scripts and SQL commands. For the analysis of

the extracted queries, the following key statistical techniques were availed of in whole or in part, albeit some of the outputs having been retracted in an effort to facilitate the ease of comprehension by the ordinary reader¹:

1. Descriptive and Inferential analysis
2. Time series analysis
3. Hierarchical cluster analysis
4. Support vector machines
5. Artificial Neural Networks

To achieve this task, several platforms and tools were used, with the main ones being R v3.51, Stata MP 8 Core v15, EViews v10, and IBM SPSS Modeler v18, each playing a special role. The full replication procedures can be provided, if need be, along with the syntax and replication commands.

3. Results

This section presents the key findings from the several techniques that were used. The first procedure used was the assessment of the voters' roll, hereinafter, referred to as the VR, from a demographic point of view. The analysis of the population dynamics was achieved by benchmarking the observed distributions against the estimates of the voter population, along with benchmarking against the 2013 and 2008 voters' roll data, as well as the 1982-2012 census data with the key goal being to help determine whether the two distributions were consistent or not. The second part delves into confirmatory tests to further test the outcome of the exploratory analysis findings, and these tests include time series analysis, spectral analysis and artificial neural networks, which eventually informed on the suspect extraction.

3.1 Demographic Distribution

Based on the estimates by ZESN ² as well as the VR V1.0³ the distribution of the possible voter population and the eventual voters that registered is presented below. From the analysis, the total voter population expected in 2018 was 7224129, and from the numbers extracted in the first version of the voters' roll, the total number of

¹ Key advanced procedures and outputs will be available upon request.

² ZESN (2017) 2018 Population Projections for Zimbabweans Aged 18+, ZESN, Harare

³ The initial Voters' Roll released by ZEC on the 18th of June

registrants stood at 5683936, with the estimated number of the voting population that did not register being substantially high, that is, 1540193.

Table 1: Voter Population Estimates and the Final Voters Registered

	EXPECTED			REGISTERED		
	Men	Women	Total	Men	Women	Total
Bulawayo Metropolitan	186055	223334	409389	120387	138303	258690
Harare Metropolitan	651540	694278	1345818	443178	457122	900300
Manicaland	405871	506891	912762	318219	415074	733293
Mashonaland Central	296783	323496	620279	252021	279843	531864
Mashonaland East	341616	394173	735790	289475	343651	633126
Mashonaland West	411007	423407	834414	318184	326790	644974
Masvingo	321247	433067	754314	260692	356512	617204
Matabeleland North	177596	211996	389592	148948	189903	338851
Matabeleland South	160059	196527	356586	113734	150426	264160
Midlands	397689	467496	865185	351586	409888	761474
TOTAL	3349463	3874665	7224129	2616424	3067512	5683936

Overall, women had the highest proportion, 53.97%, as compared to men, whose proportion was 46.13%. Further breaking down the above findings, we established the relative proportion of the final number of voters vis-à-vis the estimates. The respective proportions are summarized in Table 2 below.

Table 2: Proportion of Registered Voters against the Estimates⁴

	Men	Women	Total
Bulawayo Metropolitan	64.71%	61.93%	63.19%
Harare Metropolitan	68.02%	65.84%	66.90%
Manicaland	78.40%	81.89%	80.34%
Mashonaland Central	84.92%	86.51%	85.75%
Mashonaland East	84.74%	87.18%	86.05%
Mashonaland West	77.42%	77.18%	77.30%
Masvingo	81.15%	82.32%	81.82%
Matabeleland North	83.87%	89.58%	86.98%
Matabeleland South	71.06%	76.54%	74.08%
Midlands	88.41%	87.68%	88.01%
Aggregate	78.11%	79.17%	78.68%

⁴ The accuracy of the interpretations are based on the accuracy of the geometric formula $R = \frac{Pt^N}{P_0}$ used by ZESN (2017), where Pt- base population (last census), R- Is the growth rate, N = number of years after the base year

The results above generally point to an aggregate 78.68% success rate of the BVR, at least up to the time the first voters' roll was released⁵. Nonetheless, the least aggregate success rate was observed from both Bulawayo and Harare, whose BVR success ratios were 63.19% and 66.90% respectively. The least success rate of the BVR for the registration of men was observed in the Bulawayo Metropolitan (64.71%) with the second least being Harare Metropolitan Province (68.02%). With respect to the poor registration of women, this was observed in Bulawayo Metropolitan (62.93%) while Harare was second, with a success rate of 65.84%. The third rated in both instances was Matebeleland South.

Significant success rates⁶ were, however, observed among the women category where the BVR success rate (79.17%) relatively out-surpassed those for the success of the BVR among men (78.11%). Among the men, the highest BVR success rate was observed with Midlands, being 88.41% seconded by Mashonaland Central (84.92%), with Mashonaland East being the thirds rated (84.74%). With respect to the women, the highest BVR success rate was identified with Matabeleland North (89.58%), while Midlands was the second rated, with a success rate of 87.68%, with Mashonaland Central and East being the third and fourth rated respectively. With a view to determining whether the correlation in the BVR success rates between the two groups were statistically significant or not, the Pearson correlation coefficient was computed. The respective output is presented in Table 3.

Table 3: Correlation Analysis – BVR Success Rates for Men against Women

		Men	Women
Men	Pearson Correlation	1	.956**
	Sig. (2-tailed)		.000
Women	Pearson Correlation	.956**	1
	Sig. (2-tailed)	.000	

** . Correlation is significant at the 0.01 level (2-tailed).

Based on the results, $r=0.956$; $p=0.000$. With the correlation coefficient being positive and greater than 0.85, and the p-value being less than 0.01, it can be argued that the correlation between the success rates of the BVR between males and females was

⁵ The rate has increased slightly owing to the release of the second voters' roll.

⁶ While this outcome is rather atypical, its accuracy depends on the accuracy of the estimates used. However, assuming that the estimates were accurate, then these 5 provinces with extraordinarily high proportions of male registrants would have to be flagged for further investigation.

very high, and significant. Nevertheless, the registrations in the Bulawayo Metropolitan and Harare Metropolitan were identified with very poor success rates.

3.2 External Inconsistency as Evidence of Potential Manipulation

Beyond the assessment of the estimates, we investigated the relative level of external consistency of the distribution of the age of voters. This evaluation was done in three tiers. The first tier was to gauge the observed distribution against the past population distributions and the population data was obtained from ZIMSTAT.⁷ The second tier was achieved by comparing the most recent voters' roll with the 2013 voters' roll. The respective summary statistics are presented in Table 4.

Table 4: Age Distribution, 1982-2012, Including 2013 and 2018 Voters' Rolls

Age	1982	1992	2002	2012	2013VR ⁸	2018VR	% Δ ⁹	Excess
<20 ¹⁰	804332	1252324	1509345	1415030	45541	213866	369.61%	-
20-24	655843	993143	1227627	1198202	223891	776590	246.86%	-
25-29	525496	714543	991575	1134093	546133	729749	33.62%	-
30-34	392921	608355	732919	922701	880476	787786	-10.53%	92690
35-39	318706	490518	506434	738305	899288	743248	-17.35%	156040
40-44	282130	364967	436111	525900	760980	625281	-17.83%	135699
45-49	227324	289823	357980	348753	482267	480794	-0.31% ¹¹	1473
50-54	203060	281519	302430	351600	463313	290647	-37.27%	172666
55-59	128451	182034	211728	282943	392246	288688	-26.40%	103558
60-64	142390	180310	194611	227197	297446	256332	-13.82%	41114
65-69	77822	102438	132659	167468	218029	187589	-13.96%	30440
70-74	60427	121153	123246	137989	193386	124566	-35.59%	68820
75+	86337	120825	164898	227376	471118	178800	-62.05%	292318
Total	7501470¹²	10412548	11631657	13061239	5874114	5683936	-3.24%	1094818

From the findings, the relative high proportion of the youthful age-groups is evident, with the 15-19 age group being the highly represented, followed by the 20-24-year age category, the 25-29 year-age categories, et cetera, and the consistency is clearly evident from 1982 up to 2012, as presented in Figure 1 below.

⁷ ZIMSTAT (2015) Population Projections Thematic Report. ZIMSTAT. Harare

⁸ VR – Voters' rolls

⁹ Percentage change, calculated comparing the 2013 and 2018 voters' roll data

¹⁰ NB: The Voters' Rolls only comprised of those greater than 18 years

¹¹ Anomaly 1

¹² Total population for everyone, including minors for 1982, 1992, 2002 and 2012. The voters' roll totals only reflect the 18+ year totals.

Despite the high frequency of the youthful age groups in Table 4¹³, from the illustration below, the poor representation of the younger age groups on the VR is shown. Benchmarking the 2018 Voters' Roll against the 2012 voters' roll distribution, from the intersection, it is evident that those below the 35-39-year category were under-represented. Using the 2013 voters' roll distribution against the 2012 distribution, it is also clear that those below the 30-34 years age category were under-represented.

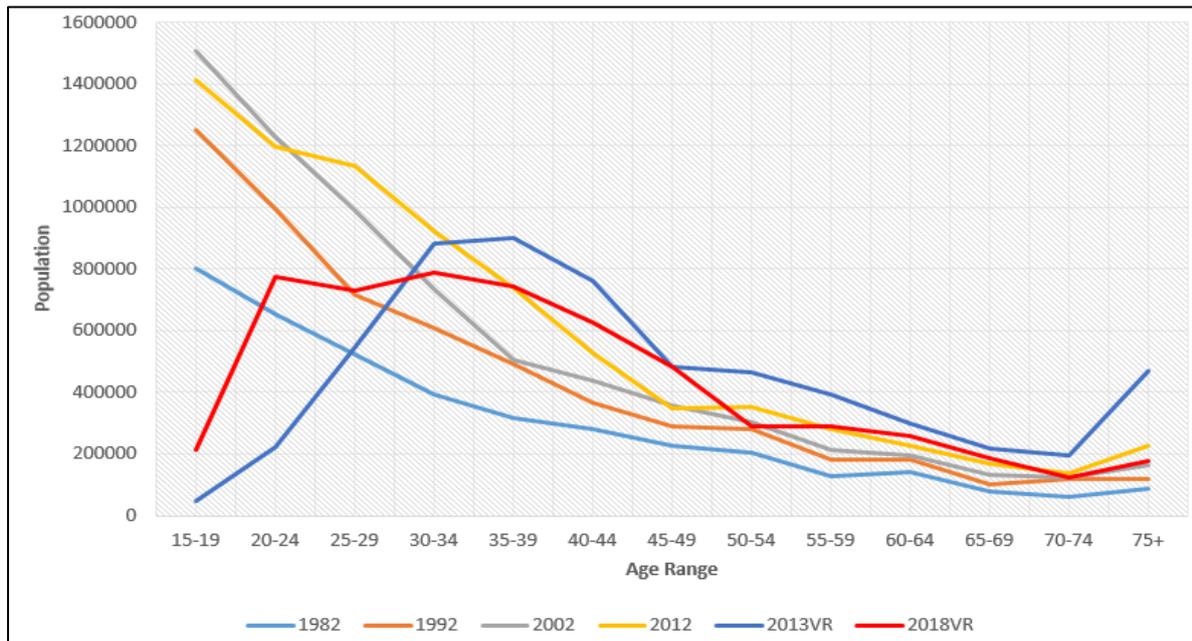


Figure 1: Age Group Distribution

Comparing the 2013 and 2018 voters' rolls, in 2018, there was 369.61% increase in the number of registrants aged less than 20 years. Further, there was an increase of 246.86% in the voters aged between 20 and 24 years in the 2018 voting season as compared with the 2013 voting season. With respect to those aged between 25-29 years, the increase in registrants was 33.62%. Nevertheless, it should be noted that from 30 years onwards, there was a significant negative change with less numbers being observed in 2018 than in 2013. Thus, mapping the two distributions, the most rudimentary minimum estimate of the number of voters who were dead by 2013 was at least 1094818. This phenomenon is further explained by the time series in Figure 2, and a more accurate computation of ghost entries in the 2013 voters' roll is presented in section §3.2.1 below.

¹³ Based on census results.

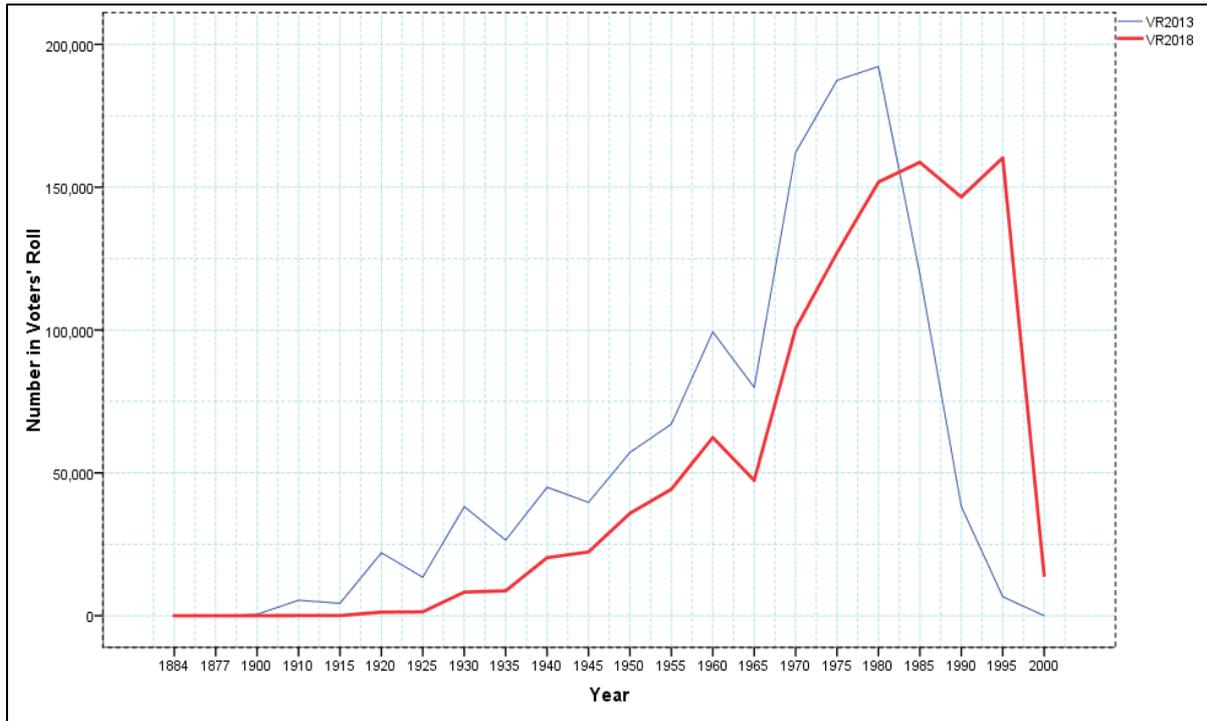


Figure 2: Age Distribution in the 2013 and 2018 Voters' Rolls¹⁴

From the foregoing, considering the 2013 voters' roll, the under representation of young voters born after 1982 is evident. However, the excess in the number of voters born before 1983 is again evidence of the presence of an unexpected number of registrants if we benchmark the 2013 voters' roll against the current 2018 voters' roll. Technically, this is evidence that there were voters in the 2013 voters' roll that we cannot account for, and these seem likely to be deceased people, but they may also reflect previous attempts at inflating the roll, to the rate of least a million voters. Further factoring in the population growth in the last five years, and other factors such as voter registration apathy among the youthful voters have an additive effect towards the cumulative total of the deceased entries¹⁵ on the 2013 voters' roll.

The phenomenon of deceased entries in the 2013 voters' roll is evident from the discrepancies in the distributions of the older voters, as summarized in Table 5 below. Disregarding the outliers, and possibly wrong entries for 1877 and 1884 for the 2018 voters roll, comparing the number of entries born in 1900, on the 2013 voters' roll, they

¹⁴ Anomaly 2

¹⁵ Possibilities of human manipulation cannot be ruled out, though the word deceased is being used as an umbrella euphemism for all illegal entries in the voters' roll.

are 484 against the 2018 voters' roll of 3. With respect to those born in 1907, in the 2013 voters' roll, they are 1000 entries as compared with the 2018 voters' roll which has only 6 entries. The same trend can be seen for all the other ages, and from Figure 1, the number of entries per age-group in the 2013 voters' roll exceeded the population census results done in the previous 2012 year, at least from voters' roll entries that were 30 years old onwards.

Table 5: Distribution of the Oldest Entries in the 2013 and 2018 Voters' Rolls

Year	2013VR	2018VR
1877	0	1 ¹⁶
1884	0	1 ¹⁷
1899	0	1
1900	484	3
1901	732	9
1902	825	4
1903	558	4
1904	838	7
1905	979	11
1906	1007	7
1907	1000	6
1908	1502	18
1909	2569	18

Year	2013VR	2018VR
1910	5412	46
1911	3596	25
1912	4321	45
1913	2515	35
1914	6388	85
1915	4343	70
1916	5362	122
1917	5399	92
1918	21648	534
1919	11398	389
1920	22017	1240
1921	12978	831
1922	15639	1139

Year	2013VR	2018VR
1923	9298	776
1924	14605	1399
1925	13425	1356
1926	12245	1399
1927	15321	2001
1928	19088	2754
1929	22694	3750
1930	38199	8262
1931	18892	4531
1932	36051	9928
1933	17808	5059
1934	21965	6971
1935	26446	8697

All this evidence culminates in the argument that the 2013 voters' roll was tainted to a great extent by deceased voters and this is a clear indication that the 2018 BVR exercise has indeed successfully helped towards cleaning up the once-in-shambles, 2013 voters' roll.

3.2.1 Estimating Ghosts in the 2013 Voters Roll

In light of the evidence of high likelihood of deceased entries in the 2013 voters' roll, it was imperative to estimate the most accurate number. Initial rough estimates showed that these were at least 1,094,818. However, to better model the magnitude of the ghosts, it was important to consider the modelling of the two distributions up to the point of intersection, which was in 1983. The resultant models are illustrated in Figure 3 below. The corresponding distributions for the two are:

¹⁶ This entry was not in the 2013 voters' roll either by ID or by age. According to international trends, it would be impossible for the individual to be alive, and hence was, most likely, a data capturing error.

¹⁷ This entry was not in the 2013 voters' roll either by ID or by age.

$$y_{2013} = 2497.54 - 95.33x + 16.74x^2 + 0.15x^3$$

$$y_{2018} = -2075.60 + 296.89x - 13.14x^2 + 0.39x^3$$

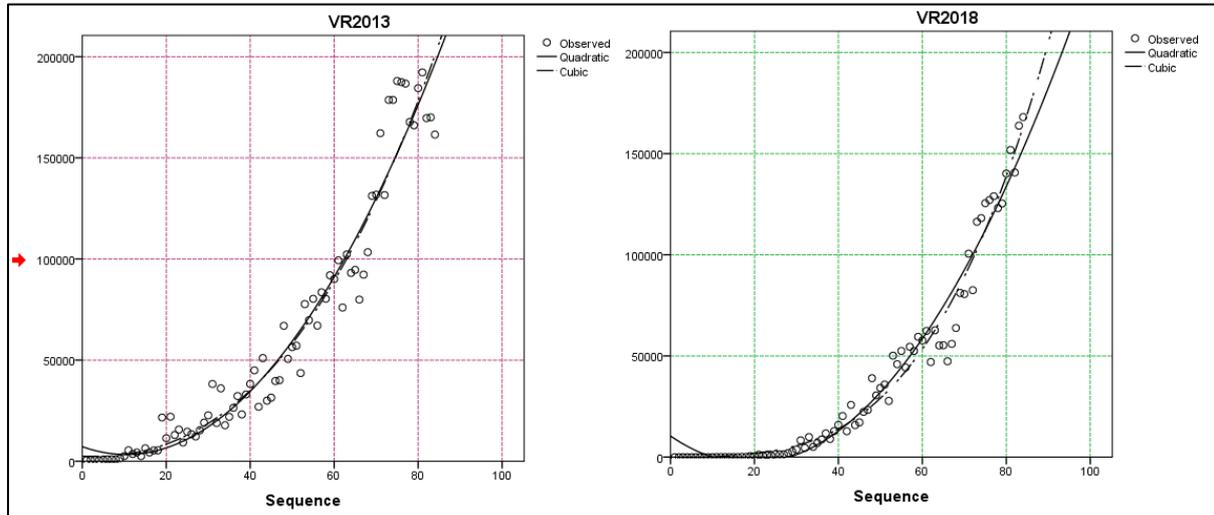


Figure 3: Population Distribution Models for VR2013 and VR2018

It follows then that the distribution of the ages in the voters’ roll followed a cubic function, where the respective r-square statistics for the two were:

Table 6: Cubic Function Model Fit

	R	R Square	Adjusted R Square	Std. Error of the Estimate
VR2013	.976	.953	.951	13357.077
VR2018	.988	.976	.975	7434.576

Based on the output above, the goodness of fit of the two VR age distribution functions was significant, with a high r-square statistic of 0.953 for VR2013 {F(3,80)=540.106, p<0.01}, and an r-square of 0.988 for VR2018 {F(3,80)=1090.395, p<0.01}. To this effect, the above equations explained 95.3% and 98.8% of the variation of the 2013 and 2018 age distributions respectively, thus confirming the accuracy of the models.

Effectively, the computation of the magnitude of the gap was considered as being affected by two factors. First was that the best estimate of the deceased voters according to today’s standards was $d \in VR$ where d was the complement of

¹⁸ Where x is the number of years from 1990. Dates beyond 1990 were trimmed being outliers

$VR_{2013} \cap VR_{2018}$. The second factor was that the result would be an underestimate, given the population growth between 2013 and 2018. Effectively, a weighting function would be imperative to account of the population growth. According to today's standards, the best estimate of the number of deceased voters (d) would, therefore, be:

$$d = \int_1^n y_{2013} dx - \int_1^n y_{2018} dx$$

where

$$y_{2013} = \int_1^{86} (2497.54 - 95.33x + 16.74x^2 + 0.15x^3) dx$$

$$\therefore y_{2013} = 5,443,639.21$$

$$y_{2018} = \int_1^{86} (-2075.60 + 296.89x - 13.14x^2 + 0.39x^3) dx$$

$$\therefore y_{2018} = 3,514,350.30$$

thus

$$d = 5,443,639.21 - 3,514,350.30$$

$$d = 1,929,289$$

19

From the foregoing computation, benchmarking on the 2018 voters' roll model, it follows then that the best estimate of the deceased voters in the 2013 voters' roll was 1,929,289. Further factoring in the population growth rate²⁰, to backtrack the 2018 model to 2013, and adding the difference to the above total, it then meant that just above 2 million of the entries in the 2013 voters' roll were deceased voters. Gauging by today's standards, it then meant that the number of bona fide voters in the 2013 voters' roll was at most 3944825, and not anywhere close to 5.8 million voters²¹.

3.2.2 Exploring the Possible Evidence of Manipulation of the 2018 VR

In spite of the positive change that could be drawn from the BVR exercise, that is, the elimination of illegal entries in the voters' roll, two key anomalies that could still point

¹⁹ NB: While the expression above is correct to 2dp, the computation was not done at 2dp

²⁰ $R = Pt^*(1+r)^n$ as used by ZESN (2017), where Pt- base population (last census), R- Is the growth rate, N = number of years after the base year

²¹ These findings tend to resonate with the findings by RAU (2014) Numbers out of Tune? An examination of the vote in Hamonised July 2013 Election. Governance Programme, Research and Advocacy Unit. Harare.

to the potential manipulation of the voters' roll emerged from the distribution of the age data were observed. First, considering Table 4, the least deviation between the 2013 voters' roll and the 2018 voters was observed in the 45-49-year-old category, and this was a paltry-0.31% vis-à-vis other age groups which were in the excess of 10+%.

Second, considering the distribution in Figure 2, despite the fact that just above 2 million deceased entries were, inevitably, evident in the 2013 voters' role, removing these off the scene²² would affect the population distribution of the two time series viz. the 2013 and 2018 voters' roll distributions. This understanding is founded on the basis of the varying death rates that inevitably change depending with age and other socio-economic factors and can be exemplified with the different trends between 1900 and 1925. However, the atypical homogeneity of the time series between 1925 and 1985 is uncharacteristically unusual, let alone the characteristic depression between 1960 and 1970, as well as the steep increase between 1965 and 1975. To better illustrate the discrepancy, the comparison of the official age pyramids is ideal. The corresponding age distribution pyramid for the Zimbabwean population based on the 2012 data by ZIMSTAT is presented in Figure 4 below.

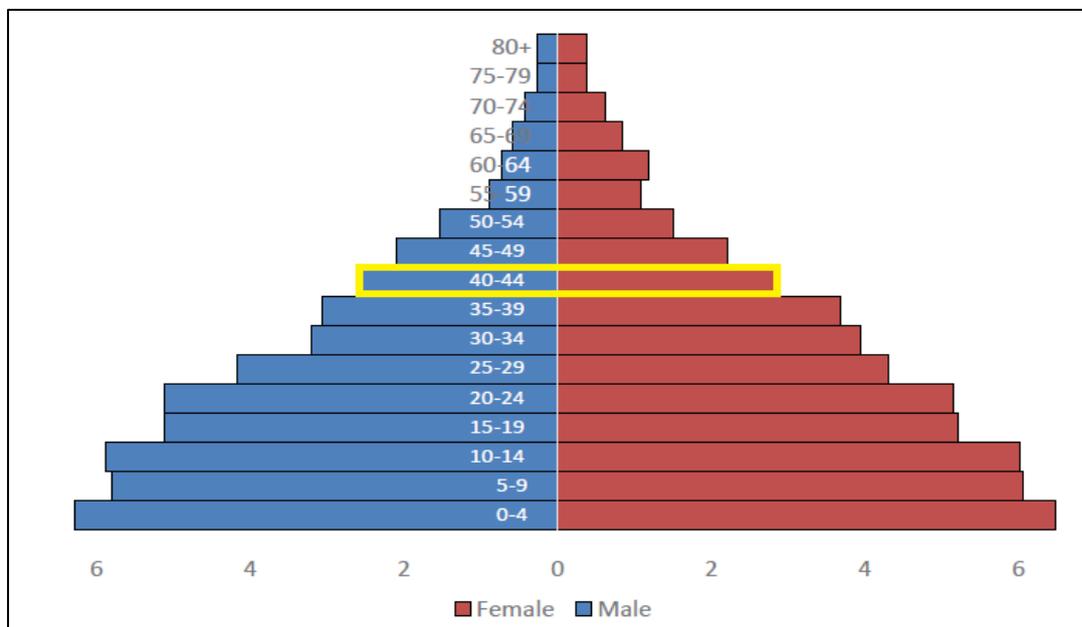


Figure 4: The 2012 Population Pyramid of Zimbabwe, Zimbabwe, 2012 Census²³

²² As expected from the BVR exercise

²³ ZIMSTAT (2015) Population Projections Thematic Report. ZIMSTAT. Harare

The 40-44-year age group in 2012 above is now the 45-49-year age group in the 2018 voters' roll as shown in Figure 4 overleaf. Considering the fact that the sample that registered to vote was 5,683,936, off an approximate maximum of 7,224,129. This accounts of 78.68%, the number in the voters' roll is a significant number that is adequate to replicate the typical true-to-life age pyramid. Further considering power analysis²⁴, for a population of 7,224,129, at 1% margin of error, 99% confidence level, and 78.68% registration rate, the minimum adequate sample for the replication, or rather, representativeness of the population parameters is 11,113. In our case, we had 5,683,936 entries, which is 511 times more than the sample required. To this effect, the sample considered, that is, the 2018 voters' roll entries were supposed to recreate a true-to-life distribution of the age distributions. However, this failed, as shown below.

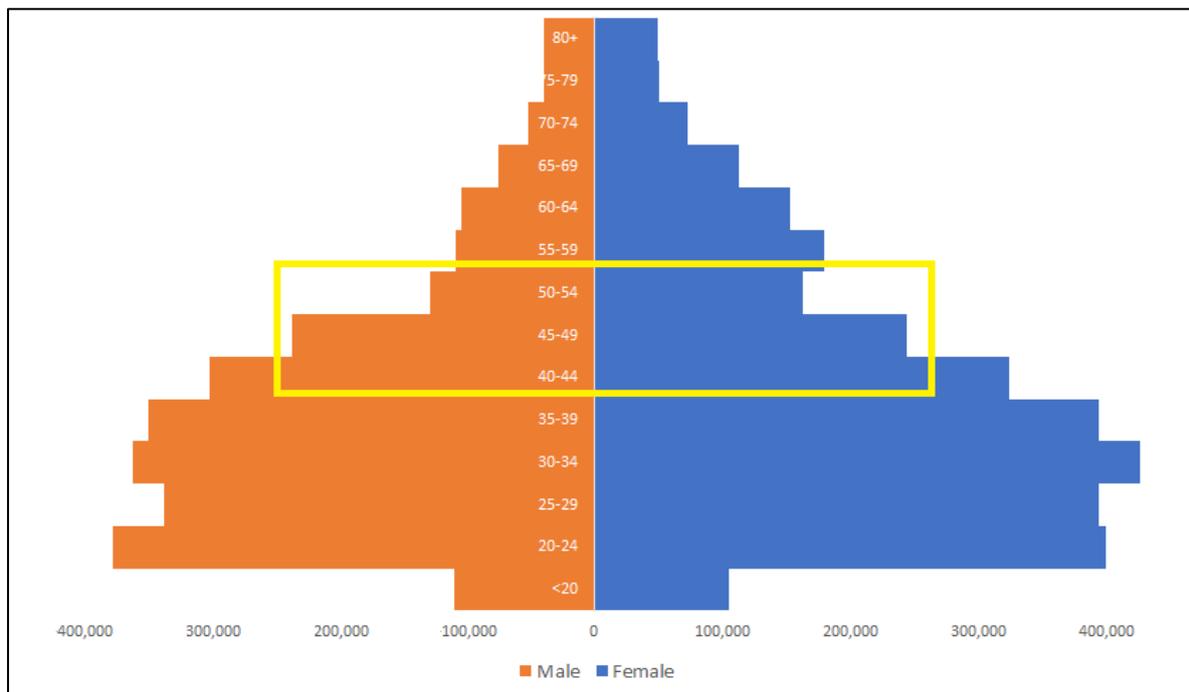


Figure 5: Population Pyramid – 2018 Voters' Roll Entries

From the foregoing, there was a significant deformation in the pyramid arms between the 45-49-year-old category and the 50-54-year-old category. The bottom deformation in Figure 5 from 30 years and below, can best be explained by the poor voter registration levels by the youthful ages as shown in Figure 1, as well as the blue highlight in Figure 6 below.

²⁴ $n = N_x / ((N-1)E^2 + x)$; where $x = Z(c/100)^2 r(100-r)$, $E = \text{Sqrt}[(N- n)x/n(N-1)]$

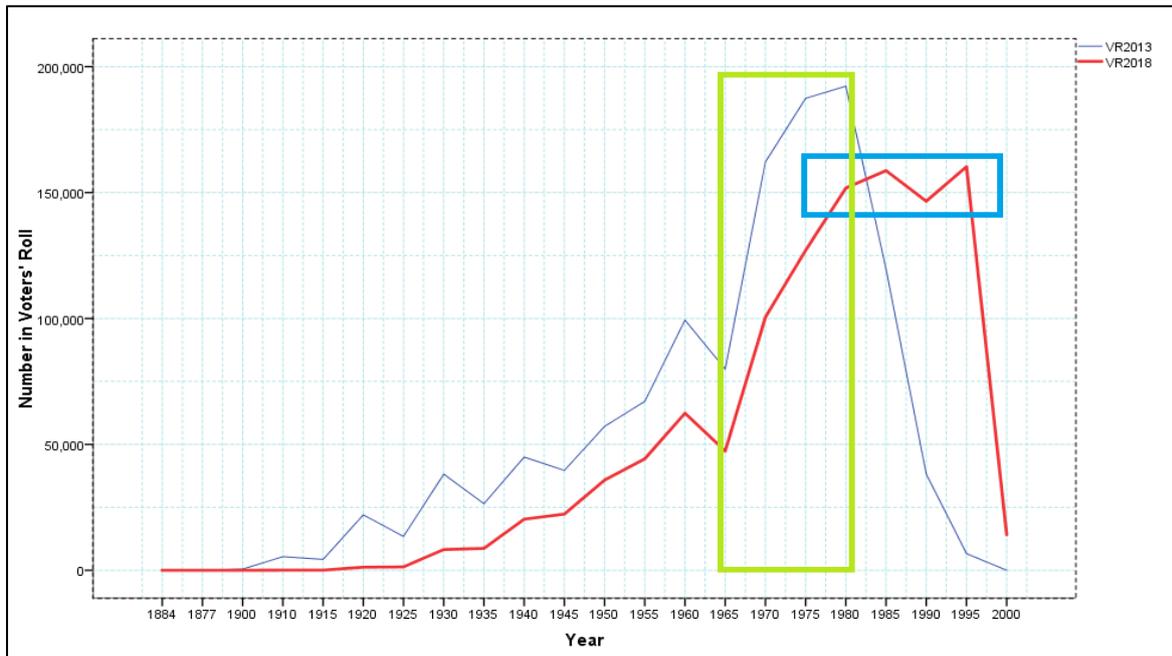


Figure 6: Age Distribution in the 2013 and 2018 Voters' Rolls²⁵

However, the steep gap between the 45-49-year age category is characteristic in that the gap between the ages is bigger than the estimates, along with the steep increase as shown by people in the voters' roll that were born between 1965 and 1975, as shown by the lime highlight above. Comparing with the projections by ZIMSTAT:

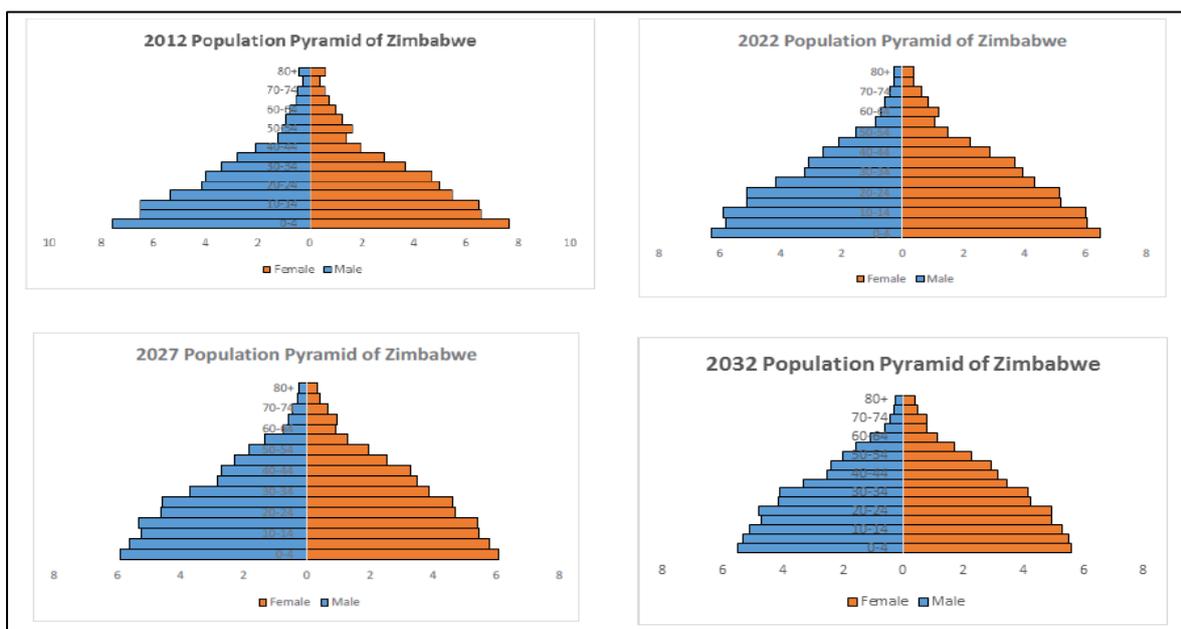


Figure 7: Population Pyramid Projection Profiles: 2012 – 2032²⁶

²⁵ Anomaly 2

²⁶ ZIMSTAT (2015) Population Projections Thematic Report. ZIMSTAT. Harare

From this basis, there are three possibilities. First, it is possible that the methodology used to come up with the estimates By ZIMSTAT were possibly wrong, which we cannot dismiss, but consider as highly unlikely, with their models having been confirmed by UNFPA estimates. Second, in the case where we can safely assume the projections by ZIMSTAT to be accurate, which was our best shot, this points to the voters' roll as being inconsistent with the norm. Then third, the possibilities of tampering with the voters' roll may thus not be ruled out, in cases where either multiple registrations were done using the same or different identities, or cases where additional entries were clandestinely added onto the voters' roll, though in oblivion of the effect that such an action would have towards the distortion of the demographic distributions. This latter possibility was relatively inconceivable but remained open for further investigation.

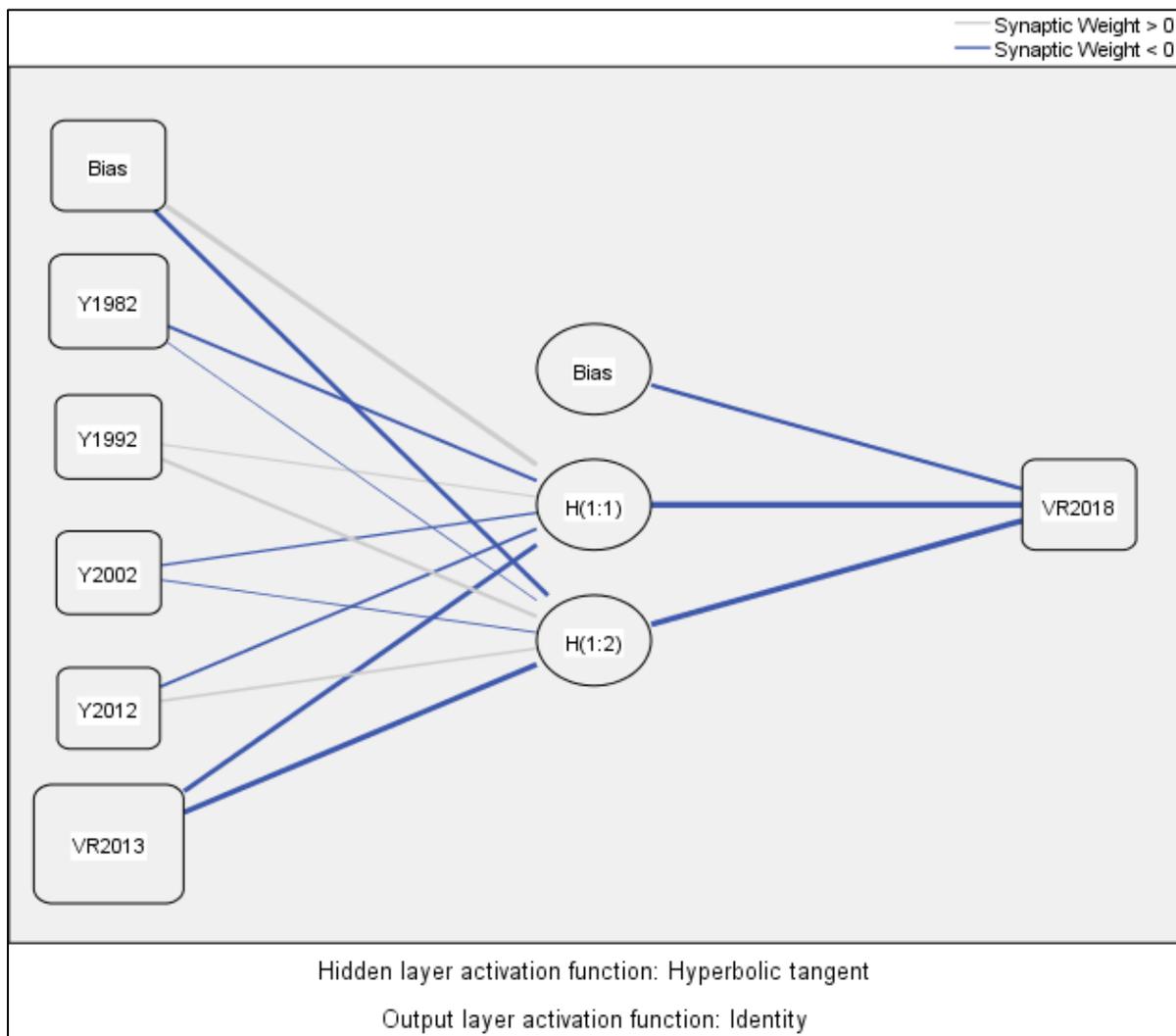


Figure 8: Artificial Neural Network Output

To further investigate the anomaly, two considerations were made, that is, advanced neural networks as well as time series analysis. With respect to the neural network approach, multilayer perceptron (MLP) network trained with backpropagation was used, with the key advantage being that it is a supervised learning algorithm. A dataset of the distribution of the 2018 ages from 18 years upwards were tested against the distribution of ages from using 5 distributions as the predictors viz. the 1982 census distribution, the 1992 census distribution, the 2002 census distribution, the 2012 census distribution and the 2013 voters' roll distribution. The findings from the analysis are presented in Figure 9 below.

It is evident from the output that the distribution of the age was less predicted by the previous 4 censuses' distributions, which is odd, but rather, by the 2013 voters' roll distribution of ages. The respective normalized importance of each of the predictors is presented in Figure 9.

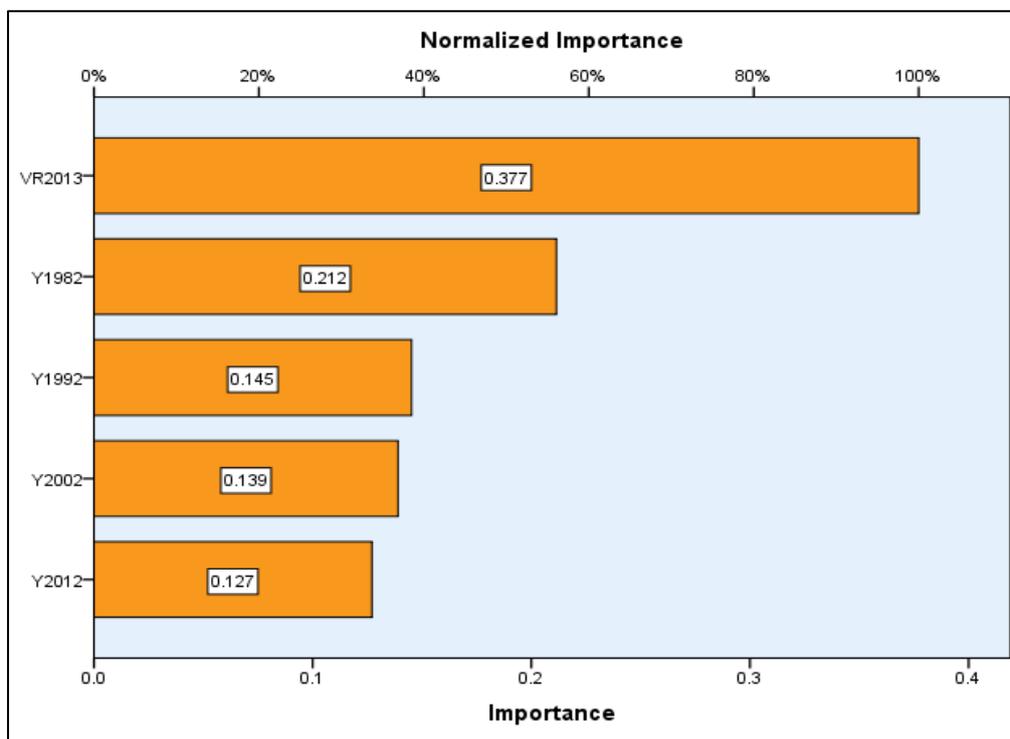


Figure 9: ANN Normalised Importance

The output above technically qualifies the 2013 voters' roll as being the main predictor of the distribution of the age distributions in the 2018 voters' roll, over the earlier distributions. This finding is rather odd, given the fact that the distribution of the

demography in the 2013 voter' roll was biased owing to the high proportion of deceased entries. While this finding does not ordinarily qualify the latter 2018 dataset as being to some extent, a derivative of the 2013 dataset, the findings tend to qualify that there were significant anomalies in the distribution of the ages between the 2018 voters' roll vis-à-vis the trends in the previous censuses²⁷.

With a view to further evaluating whether there was a structural break or not, breakpoint unit root tests were conducted based on the Dickey-Fuller autoregressive coefficients with a maximum lag of 11 using the Schwarz information criterion. The respective output is presented in the Figure 9 below. A structural break in the time series was found to exist at the 1968 mark, and thus the separation of the entries in the voters' roll before and after 1968.

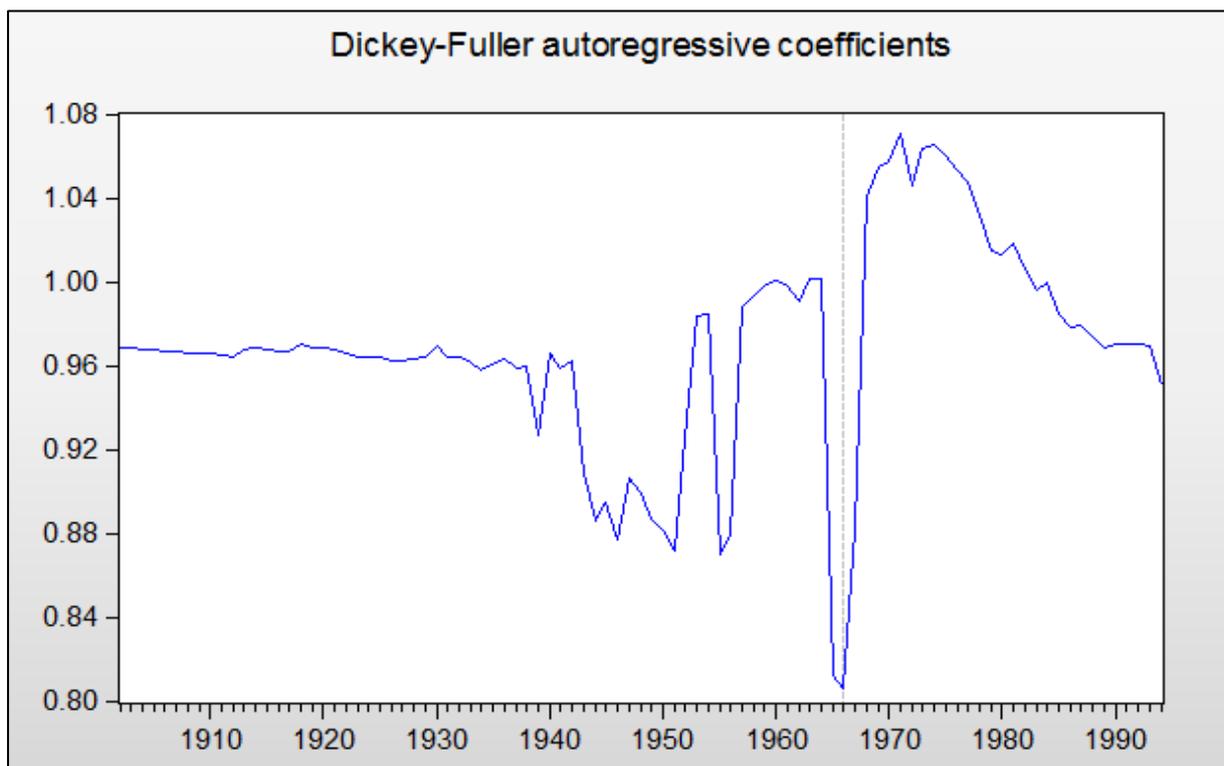


Figure 10: Dickey-Fuller Autoregressive Coefficients²⁸

Subtracting 1968 from 2018, gives 50, and this resonates with the earlier findings. The structural break validated above tends to validate the unique structural break observed

²⁷ While tempering cannot be ruled out, what is key to note that this evaluation is not definitive, but rather, it is going to be through a rigorous case-by-case audit that we can rule out the possibility.

²⁸ The full tests carried out can be provided, if need be.

in Figure 4 earlier between the 45-49 age group and the 50-54 age group. These anomalies were more characteristic in both the 2013 and 2018 voters' roll distributions, but not peculiar in the distributions of the demographics from the census records. Also comparing with the population growth rates between 1960 and 1975 in Figure 11 below, does not show any significant deviation that could explain the discrepancy in Figure 2 and Figure 5.

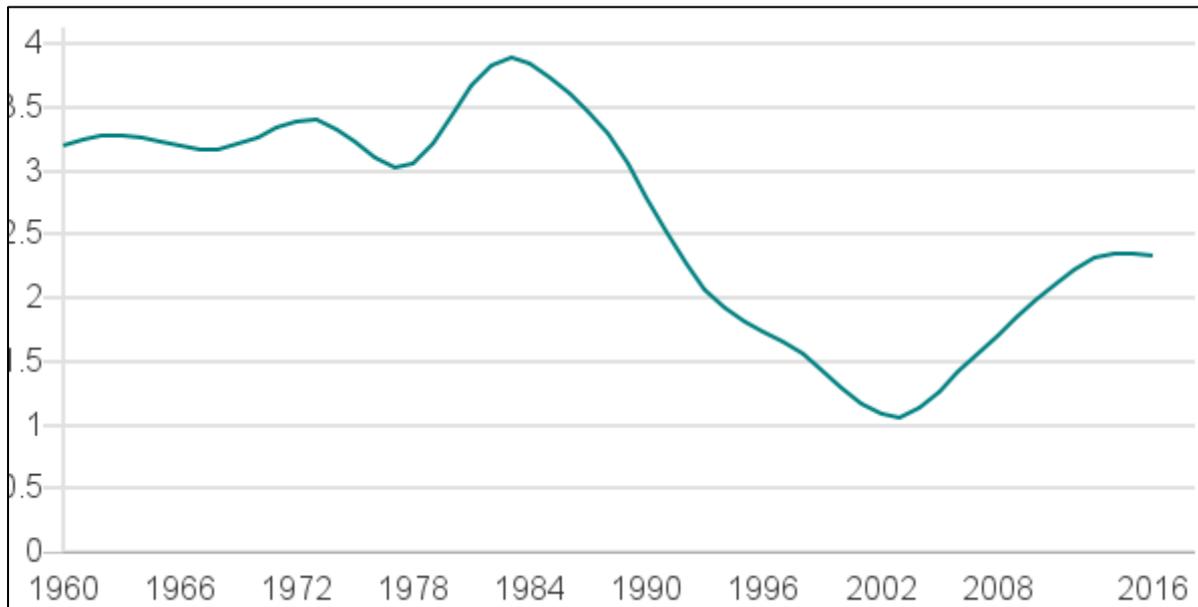


Figure 11: Zimbabwe Population Growth Rates 1960-2016²⁹

This phenomenon where there is an asynchronous structural break, which is undocumented, at least from the research done, could be possible evidence of deliberate addition of entries onto the voters' roll, with the bulk of these being stuffed being persons from the early 70s. To further confirm or disconfirm this argument, a total of 403,095 and 403,107 pair-wise anomalies between the 2013 and 2018 voters' roll entries from 1900 to 2000 were extracted and analysed to help establish their distribution pattern. The corresponding time series is presented in Figures 12 and 13.

From the distributions, it is evident that the modal YOB (year of birth) for the anomalies was 1972, while a key point to note was that there was a significant drop in the distribution in 1965 as shown in Figure 12. Referring back to the population pyramids, this outcome conforms with the distributions.

²⁹ Source: World Bank (2016)

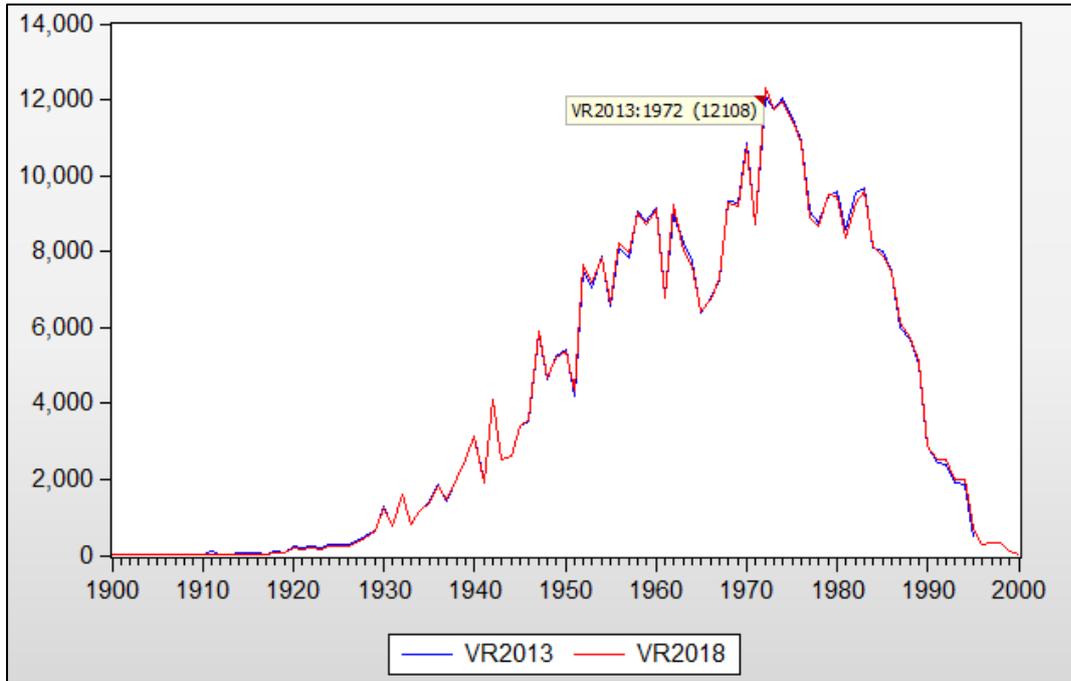


Figure 12: 1972- Distribution of Flagged Anomalies (2013VR vs 2018VR)

Converting to age, one born in 1972 would be 46 years this year. The significant detrending of the anomalies is seen before 1972, with a significant drop in 1965.

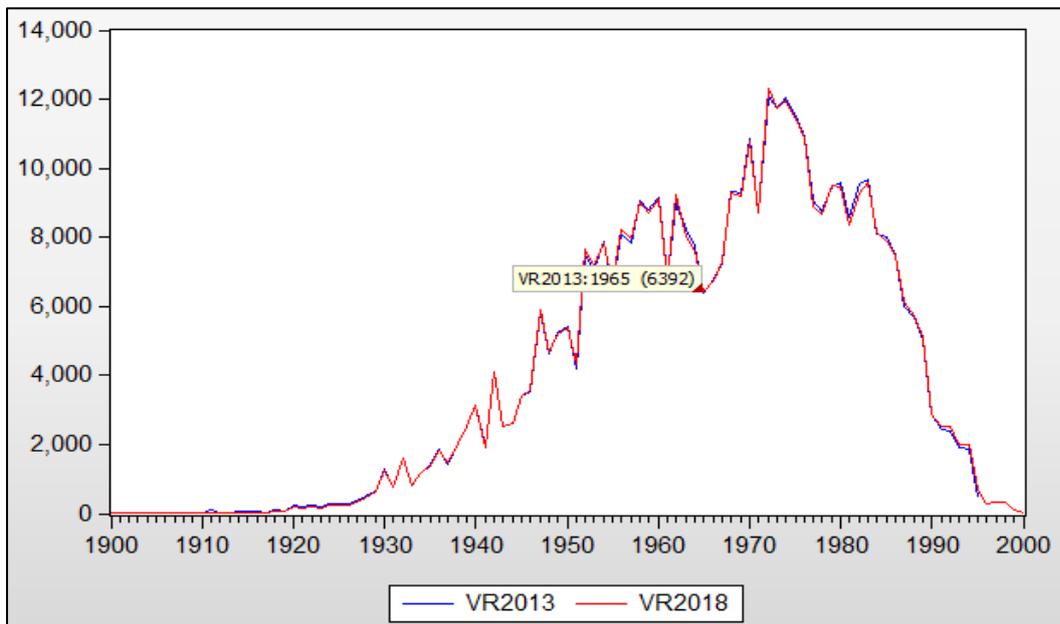


Figure 13: 1965- Distribution of Flagged Anomalies (2013VR vs 2018VR)

From the above, converting the YOB 1965 to age would translate to 53. In this regard, the age pyramid distribution in Figure 4 is confirmed, the key point of argument being the unexplained structural break between the 45-49-year category and the 50-54-year

category. However, while this argument does not necessarily relate the 2018 voters' roll as a derivative of the 2013 voters' roll, what is clear are the relative synchronous relationships between the two distributions. It would thus not be technically incorrect to assume or propose that the 2018 voters' roll could have been manipulated at a greater scale the modal YOB range of these entries being within the 1965-1980 period.

To further support the possibility of human manipulation, the demographic distribution for each and every province was computed. From Figure 14 below, four distributions tend to vary to some extent with the others and these correspond to Harare, Bulawayo, Matebeleland North and Matebeleland South. Nevertheless, there tends to a common synchronous trend with respect to the rest of the provinces, that is, all Mashonaland Provinces, Midlands, Masvingo and Manicaland.

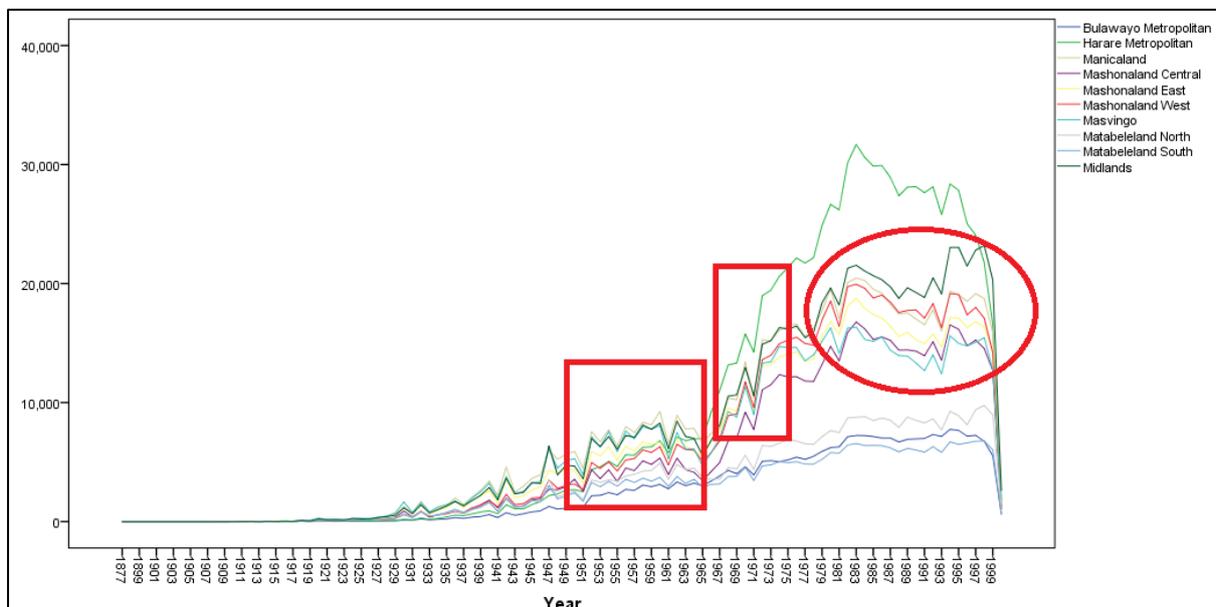


Figure 14: 1965- Distribution 2018 Ages by Province

The synchronous pattern in the population distribution models above tends to defy logic. One key reason to note is that the population distribution trends are never the same for various provinces³⁰, along with the dynamics that affect population growth such as mortality, mobility rates, among others as put forward in the ZIMSTAT's Zimbabwe Population Census 2012³¹. A case in point are the net migration rates, which in the long run, affect the population for the specific region. The extract from the census report is presented in Table 7 below.

³⁰ ZIMSTAT (2015) Population Projections Thematic Report. ZIMSTAT. Harare

³¹ Source: ZIMSTAT (2012) Zimbabwe population Census 2012 Report. ZIMSTAT. Harare

Table 7: Provincial Net Migration Rates³²

Province	2012
Bulawayo	4.18
Manicaland	-7.21
Mashonaland Central	-4.85
Mashonaland East	3.74
Mashonaland West	6.5
Matebeleland North	-0.39
Matebeleland South	8.52
Midlands	-3.5
Masvingo	-5.1
Harare	5.12

Basing on the above net migration rates, above, the migration rates are so diverse and vary in magnitude and direction over time. Factoring in other aspects such as mortality and fertility, which also vary from province to province, there would be nowhere we could ever imagine having the same distribution in Figure 14 that was characteristic of the distribution in the 2018 voters' roll. Further referencing ZIMSTAT³³, Manicaland was projected to having an annual growth rate of 2.4%, Mashonaland Central of 2.2%, Mashonaland East of 2.2%, Mashonaland West of 2.3%, Midlands 2.1% and Masvingo 2.1%³⁴. From these findings, the growth rates differ from province to province and in this regard, the synchronous distribution observed in Figure 14 fails to meet the external validity test as it reflects the population growth rates which differ from province to province. In other words, the synchronous series observed with five of the provinces are highly unlikely and atypical, and this could be evidence to potential human manipulation of the voters' roll.

To this effect, one of the key arguments from these findings was that a unidimensional review or audit of the 2018 voters' roll just by virtue of the use of the 2018 BVR was not going to be sufficient. Rather, a multidimensional audit or rather, a retrospective evaluation of the 2018 voters' roll against principally the preceding voters' rolls was

³² ZIMSTAT (2012) Zimbabwe population Census 2012 Report. ZIMSTAT. Harare

³³ The work by ZIMSTAT is very extensive, credible and elaborate, the reason for referencing them

³⁴ ZIMSTAT (2015) Population Projections Thematic Report. ZIMSTAT. Harare

imperative. Despite the availability of the 2008 voters' roll, this was, however, not considered for the neural network training on the grounds that the level of inconsistency within the roll was very high, after several exploratory evaluations had been done to this effect. However, this was used to cross-confirm some entries that could not be confirmed by the 2013 voters' roll alone. The following section shall present the key insights that were derived from the various perspectives considered for the audit.

3.3 Internal Inconsistency as Evidence of Potential Manipulation

The preceding inquiry took a holistic demographic-based approach, and the presence of demographic anomalies did point to possibilities of tempering as the demographic distributions in the 2018 were atypical³⁵. With a view to empirically test whether manipulation was actually done, it should be stressed that the only fool-proof method to determine conclusively that manipulation had been done was the AFIS deduplication system as it would be impossible to create unique biometric attributes on a large scale. Nevertheless, despite the fact that access to the AFIS system used by ZEC would be practically impossible, this technical evaluation could equally benefit from the evaluation of the internal consistency of the voters' roll.

In light of the fact that the data available was semi-structured, with dates and regions being well defined, what was key to consider was the fact that part of the data was unstructured, and this included names, surnames and addresses. To this effect, supervised feature learning algorithms were used to help automate the classification process, which in the long run, helped save time sifting through more than 16 million records³⁶ for any inconsistencies. In total, 403,107 cases were flagged in the 2018 voters' roll based on the techniques used³⁷.

³⁵ The basis of the argument is based on the principles of power analysis (Murphy, Myos, and Wolach, 2014). Typically, we would expect a representative sample to exhibit the same characteristics as the population.

³⁶ The records used were the 2008, 2013 and 2018 voters' rolls. Despite having access to the previous voters' rolls, prediction and classification modeling was done solely based on the 2013 voters' roll, which was somewhat cleaner than the 2008 roll, the latter which was fraught with significant anomalies.

³⁷ This figure is the total of the inconsistencies that could be extracted. The inconsistencies include, but are not limited to conflicting ID entries, duplicate IDs, and other slight anomalies such as different DOBs, gender mis-assignment, marginal-to-gross difference in names among others.

3.3.1 CASE 1: Quasi-Duplicate Cases in 2018, Single Trace from 2013 and 2008

One key possibility of tempering was observed from the fact that based on the quasi-duplicate cases (based on the part of the ID)³⁸, only one of each of the pairs had a trail in the previous voters' rolls. A case in point:

2018	TSHABANGU	GRACE	08-466622-Z28	Female	3/15/1964	Umzingwane RDC
2018	NDLOVU	LUCIA	08-466622-Z56	Female	6/18/1939	Bulilima RDC

2008 NDLOVU LUCIA F 08-466622-Z-56 18-Jun-1939
 2013 NDLOVU LUCIA F 08-466622-Z-56 6/18/1939

The entry of Tshabangu Grace was never in either of the 2013 and 2008 voters' rolls. With Grace having been born in 1964, why could she not have registered to vote all along that she finally decided to vote in 2018? This is a valid question because the voters' rolls preceding the 2018 roll were a cumulative effort and had served many other elections with new entries being added and deceased entries being removed. This case where we have an aged individual emerging from nowhere, sharing an identical ID number with an individual with an audit trail deserves to be flagged and further investigation would need to be done. This phenomenon was inherently generic among the duplicate entries. Another case in point:

2018	GARUDZO	MURANDA	14-052831-Q83	Male	8/12/1948	Kadoma Municipality
2018	MAPIMELE	KETSIA	14-052831-Q22	Female	10/15/1945	Masvingo RDC

2013 GARUDZO MURANDA M 14-052831-Q-83 8/12/1948
 2008 GARUDZO MURANDA M 14-052831-Q-83 12-Aug-1948

In the above case, we have two individuals, one Mr. Garudzo and another Ms. Mapimele. Both reside in different provinces, but they have one thing in common, that is, they collected their IDs from the district, and their IDs are similar. The only difference, disregarding the suffix, which does not serve much use as far as distinguishing IDs is concerned, is the fact that one of them has an audit trail and another does not. The fact that in virtually all the duplicate IDs sharing the same

³⁸ Based on the MOD23 technique, what generally forms the ID are the numbers leading to the check letter. The suffix is in no way a unique identifier. In this regard, all the numbers leading to the check letter ought to be different, and by no means should they be the same if a unique identity is to be retained.

authentication numbers, one of them could be tracked and another could not.³⁹ These cases could only be dismissed following an intensive audit of the voters' roll and a case-by-case inquiry to ensure the internal consistency is retained.

The same trend is evident even in the youthful entries, such as that of Mr. Tomu.

2018	TOMU	ARTWELL	15-171555-V15	Male	8/18/1989	Bindura RDC
2018	TOMU	ARTWELL	15-171555-V45	Male	8/18/1989	Mazowe RDC

2013 TOMU ARTWELL M 15-171555-V-45 8/18/1989

Neither of the two were found in the 2008 voter' roll, although one entry was in the 2013 voters' roll. What should be noted in the two cases above is the coincidence of two Tomus being born on the same day, collecting the IDs from the same district, and being allocated the same ID numbers, only with different districts of origin. Taking this case into consideration, one can easily dismiss that possibly upon collection of the ID, the registry district office personnel might have looked up whether Tomu had taken an ID and assigned the same. However, in such a case, we would have seen a case of a complete duplicate, including the suffix. But, given the precedent cases, what should be noted is the blanket coincidence that despite the age, or province being voted in, for each pair of duplicates, one had a voting history and another one did not.

3.3.2 CASE 2: Change of ID Suffix from 2013 Records to Present Records

Another case of anomalies flagged was a case where individuals apparently changed their suffix between 2013 and 2018. As explained in the first report, the suffix identifies one's origin village. That said, it is close to impossible that we can change history within 5 years, at least if the suffix serves its purpose correctly.

2013	CHIKANDIWA	TENDAYI LUCIA	F	26-057660-S79	3/6/1970
	Hwange East				
2018	CHIKANDIWA	TENDAYI LUCIA	F	26-057660-S07	3/6/1970
	Gweru Urban				

In the cases above, the DOB, forenames and surname are the same, and the key part of the ID is the same. However, what is different is the suffix, and this changed from S79 in 2013 to S07. Unfortunately, none of the records were in the 2008 voters' roll.

³⁹ This is a big red flag of potential manipulation.

In light of the above, the odds are very slim that on any particular day, a baby was born to two families of the Chikandiwas (rare surname), and had the similar name and surname being the same⁴⁰,

2013	CHIKANYAWU PESEVERANCE	F	14-176496-J14	3/6/1986
	Chiredzi West			
2018	CHIKANYAWU PESEVERANCE	F	14-176496-J12	3/6/1986
	Harare North			

2008 CHIKANYAWU PESEVERANCE F 14-176496-J14; 06-Mar-1986

In the above case, the entry with the suffix J14 was in the 2008 voters' roll while the one in the 2018 voters' roll was not. The argument of typos may explain this case, in which latter case might point to the fact that the 2018 record could be the wrong record as is in the case of Ngonidzashe.

2018	PIRIKISI	NGONIDZASHE	F	26-055576-G26	1/1/1968
	Mkoba				
2013	PIRIKISI	NGONIDZASHE	F	26-055557-G26	1/1/1968
	Mkoba				

In the above case, the MOD23⁴¹ validation checked out correct in both scenarios, which is one of the weaknesses of the encryption mechanism that is used as it is not infallible. However, it is easy to dismiss the 2018 entry as wrong on two grounds. First, there is no 9-digit ID whose ID body starts with 0, and second by checking against the 2008 voters' roll. The records as retrieved were:

2008 PIRIKISI NGONIDZASHE F 26-055557-G-26; 01-Jan-1968

However, despite the overlap between typographical errors during the time of capturing, possibilities of manipulation may not need to be disregarded in light of the next scenario.

⁴⁰ Only 11.5% of the individuals in the 2013 and 2018 voter's rolls have more than one forename.

⁴¹ The check letter is simply calculated by dividing the merged numbers on the ID by 23 and then checking the remainder part of the quotient against the corresponding letter. In both cases above, the remainder was 7, and hence the corresponding check letter being G. However, I, O and U are excluded.

3.3.3 CASE 3: Change of ID Check Letter from 2013 Records to Present Records

This third case is rather intriguing owing to the significant commonalities present, and a slight change that no one but the Registrar General's office should be able to have made. In short, this presents one of the many cases where the ID completely changed, with signs of similarities with the former ID⁴². This is the case of one Ms. Chibwitiri:

2018	CHIBWITIRI NURSE MERENIYA	F	45-008618-K45	1/13/1949
	Mt Darwin North			
2013	CHIBWITIRI NURSE MERENIYA	F	45-008616-H45	1/13/1949
	Mount Darwin			

Both records check out. The name is identical, the dates of birth are identical, along with the constituency. According to the entries in both voters' rolls, Chibwitiri is a very rare surname, and having two babies with identical two forenames, being born on the same day would be technically impossible. What this means, in other words, is that these two entries are the same. The digits in the IDs are similar, though different, and the slight difference of a 6 and an 8 tells the bigger story. The fact that the check letters are different rules out the possibility of the entry of either 6 or 8 being a typo.

Testing the authenticity of both IDs⁴³, the following output was produced: For the ID 45-008618-K45, the remainder was 10, and this corresponded to the letter K⁴⁴⁴⁵ and for the ID 45-008616-H45, the remainder was 8, and hence the check letter H. In both instances, the IDs were, according to the standards of the Registrar General's Office, authentic. None of the above instances was a typo, yet we have an entry with the same name, same other credentials, reappearing in the 2018 voters' roll with a closely similar, but unique ID number.

2008 CHIBWITIRI NURSE MERENIYA F 45-008616-H45 13-Jan-1949

⁴² This is a hot red flag

⁴³ Can be easily be done in Microsoft Excel by the formula =MOD(45008618, 23)

⁴⁴ Only the fraction part of the quotient is considered: 1=A, 2=B, 3=C, 4=D, 5=E, 6=F, 7=G, 8=H, 9=J, 10=K, 11=L, 12=M, 13=N, 14=P, 15=Q, 16=R, 17=S, 18=T, 19=V, 20=W, 21=X, 22=Y, 0=Z

⁴⁵ The computation can be automated seamlessly in Microsoft Excel by the formula =IF(A1=1,"A",IF(A1=2,"B",IF(A1=3,"C",IF(A1=4,"D",IF(A5=1,"E",IF(A1=6,"F",IF(A1=7,"G",IF(A1=8,"H",IF(A1=9,"J",IF(A1=10,"K",IF(A1=11,"L",IF(A1=12,"M",IF(A1=13,"N",IF(A1=14,"P",IF(A1=15,"Q",IF(A1=16,"R",IF(A1=17,"S",IF(A1=18,"T",IF(A1=19,"V",IF(A1=20,"W",IF(A1=21,"X",IF(A1=22,"Y",IF(A1=0,"Z", "ERROR"))))))))))))))))))) otherwise, we can provide the simple program we developed if checking on a larger scale as Excel only limits rows to 1,048,576 records only. We can provide a custom SQL query, or the syntactic code for any of the statistical programs to automate the process.

Checking against the 2008 voters' roll, the 2013 ID 45-008616-H45 was in the 2008 voters' roll while the recently used ID 45-008618-K45 was not in the voters' roll. If the older ID that has been in the voters' roll for more than 2 election seasons, how could it change in 2018? This is the biggest question. Could it have been changed for a good reason? This question will be only for the Registrar General's office to answer for the many similar cases. Another similar extract from the pool worth noting is that of Mr. Murombedzi.

2013 MUROMBEDZI TAGAREPI M 70-554391-E70 1/14/1941
Zvimba West
2018 MUROMBEDZI TAGAREPI M 70-054391-B70 1/14/1941
Makonde

The 2018 case could easily be dismissed on the basis of it being a typo, where 0 was typed in lieu of 5. But, in that case, those who captured his data during the BVR exercise would be unlikely to create another typo by typing B in lieu of E. The remainder of the division of 70554391 by 23 is 5, and the check letter in 2013 is confirmed. As for 70054391, the remainder is 2 and hence the check letter B. Technically, the IDs are authentic, but how would they change within 5 years?

2008 MUROMBEDZI TAGAREPI M 70-554391-E-70; 14-Jan-1941

The record of 70-554391-E70 is in the 2008 voters' roll, while the record for the ID 70-054391-B70 used in the 2018 BVR does not have an audit trail. The change of the ID seems to be very odd and unusual, especially since the check letter has also been changed. In a somewhat similar, but different case fitting the context, is the case of Ms. Muchenje below.

26	MUCHENJE	WISE	M	05/05/1981	15-012611-15	VIL SHUTU CH NE
27	MUCHENJE	ZVISINEI SWEMA	F	01/01/1987	45-1562261T 15	VIL KUTSOKO CH NE
28	MUCHERAHOWA	EMESI	F	14/02/1980	05-000050N 68	VIL KUTSOKO CH NE
29	MUCHERECHESI	MONICA	F	20/10/1980	38-047036Y 38	VIL SHUTU CH NE
30	MUCHIRAKWENDA	INNOCENT	M	15/07/1980	63-000003M 25	VIL M'YAMPATI
31	MUDANGWE	OLIVIA	F	01/04/1980	63-000003L 14	VIL M'YAMPATI
32	MUDARA	LATMORE	M	12/11/1980	05-000005S 5	VIL SHUTU CH NE
33	MUDENJA	DEKAT	F	05/05/1981	15-012611-15	VIL SHUTU CH NE

Figure 15: Extract from the 2013 Voters' Roll – Ms. Muchenje

Comparing the 2013 and 2018 records for Ms. Muchenje:

2013 MUCHENJE ZVISINEI SWEMA F 45-1562261-T15 1/1/1987
 Mazowe Central
 2018 MUCHENJE ZVISINEI SWEMA F 15-156261-W15 1/1/1987
 Mazowe Central

The records for the 2013 voters' roll were prefixed by district 45, and yet that changed to 15 in the 2018 voters' roll. What should be noted is that in the next part, the only difference is the double 2 which was once in the 2013 voters' roll, but apparently is not in 2018, with only a single 2 present. Considering the sequence, it would be impossible for the sequence to be the same, albeit, the omission given the district codes are different. More importantly to note is the fact that the check codes are different and are all authentic. The question remains, how could the IDs change?

3.3.4 CASE 4: Change of IDs with '9s' from 2013 Records to Present Records

The fourth set of cases comprised of cases where the IDs have been changed, with a specific mention to a significant number of special cases where they were prefixed by a 9, to make the digits a 9-digit ID despite districts having 8-digit entries, save for a few such as Harare, Bulawayo, among others.

43	CHIMUFOMBO	CElestINE	F	01/11/1978	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
44	CHIMUFOMBO	CHAUYA	M	02/01/1978	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
45	CHIMUFOMBO	CHIRATIDZO	F	18/11/1992	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
46	CHIMUFOMBO	DESPERATE	F	13/11/1992	71-093785-C71	VIL CHIDZVIKOVERA HDM NYAMHONDORO CH CHIPURIRO GURUVE	020271
47	CHIMUFOMBO	DOUGLAS	M	12/01/1978	71-9086481-P71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
48	CHIMUFOMBO	DOUGLAS	M	12/01/1978	71-9086481P 71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
49	CHIMUFOMBO	EDITH	F	11/07/1978	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
50	CHIMUFOMBO	EVURICA	F	02/01/1978	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
51	CHIMUFOMBO	FELIX	M	11/11/1961	6-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
52	CHIMUFOMBO	FUNGAI	F	2/11/1978	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
53	CHIMUFOMBO	HARDLIFE	M	2/01/1978	71-093785-C71	VIL NDORO CH CHIPURIRO GURUVE	010271
54	CHIMUFOMBO	JONAS AMBROS	M	11/11/1957	71-0591516	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271
55	CHIMUFOMBO	JOSEPH	M	13/11/1978	71-093785-C71	VIL CHIMUFOMBO CH CHIPURIRO GURUVE	020271

age 6- Ward 4 021 - GURUVE_RDC Senatorial Const. GURUVE-MBIRE - 022 Parliamentary Const. GURUVE_NORTH - 065

Figure 16: Extract from the 2013 Voters' Roll – The Mysterious 9s

The corresponding 2018 record, along with a few cases of its variants are shown below.

2018 CHIMUFOMBO DOUGLAS M 71-093785-C71 1/12/1978
 Guruve North
 2013 CHIMUFOMBO DOUGLAS M 71-9086481-P71 1/12/1978
 Guruve North
 2013 NYARUGWE CECILIA F 48-9022283-J49 1/14/1960
 Mudzi North
 2018 NYARUGWE CECILIA F 48-022283-H49 1/14/1960
 Mudzi North

2013	MUPONA SHINGIRAYI	M	63-9100580-S15	1/15/1976
	Highfield West			
2018	MUPONA SHINGIRAYI	M	63-1005801-V15	1/15/1976
	Mazowe North			

The case of Douglas is the same case of Cecilia, and a bunch of cases of a similar nature, where the 2013 record was prepended with 9, and the 2018 record has dropped the 9. Only the Registrar General is likely able to explain this, given that virtually every district has similar cases, suggesting that ID numbers were altered systematically between 2013 and 2018. In all similar instances, the check letters were different, and valid, thus eliminating the possibility of data capturing errors as none of the 2018 equivalent was in the 2008 VR.

2008	MUPONA SHINGIRAYI	M	63-9100580-S15
2008	NYARUGWE CECILIA	F	48-9022283-J49
2008	CHIMUFOMBO DOUGLAS	M	71-9086481-P71

In the case of Shingirayi, the 9 was dropped, but the numbers were then appended with 1 from 80 to 801. These instances baffle the mind as to the rationale and criteria that was used. The check letter was changed and to this effect, possibilities of human error are ruled out, and only the Registrar’s Office can best have an answer to the weird phenomena, why 9-digit records that occur in previous voters’ rolls had the 9 removed, and subsequently changing the ID.

3.3.5 CASE 5: Slight Altering of IDs from 2013 Records to Present Records

The set of cases similar to the one cited below is peculiar for a few reasons. It would not be surprising that they have the same surname, Sibanda being a common surname. What could trigger one’s attention is that they do have the same name and were born on the same date, and whose probability would be very low. Further, considering the body of the ID the digits appear to have been changed slightly, which could be a coincidence. Buy the very fact that the numbers have changed, the check letter is expected to change, which is the case, but the big question is why?

2013	SIBANDA SIKANYISIWE	F	03-170952-L03	3/6/1981
	Mberengwa North			
2018	SIBANDA SIKANYISIWE	F	03-120982-W03	3/6/1981
	Mberengwa North			

Checking against prior records, the ID used during the 2018 BVR blitz, 03-120982-W03 was never in the 2008 voters' roll. Rather, the ID 03-170952-L03 was in the 2008 voters' roll. Had the check letters been similar, then we could warrant the cases as mere data capturing issues.

2008 SIBANDA SIKANYISIWE F 03-170952-L03; 06-Mar-1981

However, the fact that in 2018, there is a new ID for the same person, brings out questions as to why? The second thing to note is that ID numbers for each district are issued serially, and if in 2008, one registered to vote using the serial 170952 for the District 03, how come 10 years later, the serial of a recently issued ID is 120982? This trend was not uncommon among the flagged entries, where we have in the greater part of the flagged entries one of them having a trace, and another without. These emerging questions could best be addressed by the RG's Office.

3.3.6 CASE 6: Inconsistencies that Could be Dismissed as Data Capturing Issues

The last batch of cases comprises of those instances where the records in 2013 and 2008 did not match one's records in 2018, in what could point the 2018 BVR exercise as being fraught with errors. In short, at least 114500 records with the same ID, different names (in part due to capturing errors, or in full) were 114503. These include cases such as those below:

2013	KAKONO JOHN	M	61-006951-C61	5/16/1957
	Mount Darwin			
2018	KAKONO MADAWU	M	61-006951-C61	5/16/1956
	Mt Darwin West			
2008	No entry			
2013	Kafudzaruva Vaines	F	61-003609-V61	12/26/1958
	Zvimba East			
2018	Kapfudzaruva Vaines	F	61-003609-V61	12/26/1958
	Zvimba East			
2008	KAFUDZARUVA VAINES	F	61-003609-V-61;	26-Dec-1958
2013	KAWARE GETRUDE	F	61-005117-J61	1/5/1950
	Mount Darwin			
2018	KAWARE GERTRUDE	F	61-005117-J61	1/5/1950
	Mt Darwin East			
2008	No entry			

Cases with the same ID, same name and different dates of birth between the 2013 and 2018 voters' rolls, also checked against the 2008 voters roll were at least 123167, and cases in point include:

2018	CHIFAMBA	ALICE	F	86-063779-K71	9/26/1983
	Zvimba South				
2013	CHIFAMBA	ALICE	F	86-063779-K71	9/26/1984
	Norton				
2008	No entry				
2018	JEKE	REVAI	F	86-063837-Y86	4/14/1984
	Zvimba North				
2013	JEKE	REVAI	F	86-063837-Y86	4/15/1984
	Zvimba North				
2008	No entry				
2013	MUCHENJE	PIASON	M	86-065111-H86	3/26/1991
	Zvimba North				
2018	MUCHENJE	PIASON	M	86-065111-H86	4/26/1991
	Zvimba North				
2008	No entry				

In all the cases above, it should be noted that by virtue of the records failing to tally, that could either have been a data capturing issue when the records were first captured, or cases where the errors were made during the BVR exercise, or otherwise given the high numbers affected, the unthinkable, deliberate manipulation. With respect to Revai, for instance, an extract from the 2013 voters' roll is shown below.

41	JASI	VASCO	M	18/01/1963	2 51 7 71	WV...
42	JASI	ZIRORE	M	23/09/1970	0 7 7 70	...
43	JASU	CRENA	F	19/08/1970	0 7 7 70	...
44	JAVANGWE	LOCADIA	F	02/01/1962	11 11 16 22	...
45	JAWADU	AGNES	F	10/10/1970	18 7 79 70	...
46	JAWADU	JAFALI	M	02/02/1986	06 39 1 E 86	...
47	JAYISI	DAINA	F	02/03/1957	11 02 1 38	...
48	JECHA	CHRISTINE	F	21/11/1967	17 5 5 70	...
49	JEFREY	JANNET	F	01/01/1963	4 1 2 26	...
50	JEKE	AKISADI	M	01/11/1971	0 8 21 71	...
51	JEKE	EDWARD	M	11/06/1907	0 1 0 0 7	...
52	JEKE	REVAI	F	15/04/1984	86-063837Y 86	KILDONAN FARM MUTORASHANGA
53	JEKE	SHEBA	F	12/11/1970	09 08 50 70	...
54	JEKENASO	CHINAZA	M	02/11/1963	05 29 2 32	...
55	JEKESON	VOLONIKA	F	09/06/1963	62 1 1 63	...

Page 37- Ward 15 046 - ZVIMBA_RDC Senatorial Const. ZVIMBA - 060 Parliamentary Const. ZVIMBA_NORTH - 206

Figure 17: Extract from the 2013 Voters' Roll – Mr. Jeke

Some of the differences in the dates had marginal deviations, while others were significantly high. Given this noteworthy class of inconsistencies, along with other issues purposefully left out such as gender mis-assignment, name/surname errors,

among others⁴⁶, it should be noted that an in-depth inquiry into the cleanliness of the 2018 voters' roll is necessary.

4. Conclusion

The above is an abridged effort to establish whether the 2018 voters' roll does comprise of what could best be identified as 'ghost voters'. From the approaches used, it is imperative to note that the discrepancies uncovered amount to sufficient evidence of the existence of ghost voters. This conclusion is premised on two arguments, the first being one of the natural principles of statistics that govern sampling that the demographic distribution of the registrants in the 2018 voting season defied all logic and as it could not adequately match to the extant demographic models. The second basis lies with the absence of an audit trail for the majority of the flagged quasi-duplicate cases, which can best be argued as a deliberate attempt to manipulate the entries in the voters' roll. Other cases presented tend to buttress the argument that deliberate manipulation of the voters' roll is not dismissible. It is, therefore, recommended that in the best interest of the voters, as well as in the best interest of the credibility of ZEC, a more comprehensive cleaning of the voters' roll is imperative.

⁴⁶ These are covered in the more comprehensive technical report to be availed upon request.